SYSTEM AND METHOD FOR OBTAINING VIDEO OF MULTIPLE MOVING FIXATION POINTS WITHIN A DYNAMIC SCENE

Inventors: Takeo Kanade, Omead Amidi and Robert Collins

BACKGROUND OF INVENTION

Field of Invention

[0001] The present invention relates generally to image processing and, more particularly, to systems and methods for obtaining video of multiple moving fixation points within a dynamic scene.

Description of the Background

[0002] In video applications, it is often desirable to separately take a set of images of a spot of action, called a fixation point, with a number of cameras surrounding the fixation point. From these sets of images one can create a so-called "3D surround-view" image sequence, which will make viewers feel as if they are flying around the scenes they see. Such image sequences are also sometimes referred to as a "3D fly-around" image sequence and "spin-image" sequence. This type of display heightens the viewer's ability to perceive the 3D spatial relationships between objects in the scene.

[0003] Initial systems to produce the "3D surround-view" effect were composed of several stationary cameras pointed at a single fixation point in the scene. The most well known examples are from the motion picture "The Matrix," although several broadcast commercials have also used this technique. The drawback to this approach is that the action has to occur at a single fixation point in the scene.

[0004] Later systems for generating such effects captured the images using cameras mounted on robotic pan/tilt devices. Using a servo loop, the pan/tilt devices allow the camera to follow a moving fixation point in real-time throughout the scene. The drawback to this approach is that the system still can only fixate on one point at a time. Moreover, it is difficult to adequately compensate for the servo errors that are introduced with such a system.

[0005] Accordingly, there exists a need for a manner in which to take a video set of multiple fixation points in a dynamic scene simultaneously, such that separate 3D surround-view image sequences of those fixation points may be obtained simultaneously.

BRIEF SUMMARY OF THE INVENTION

[0006] The present invention is directed to a system and method for obtaining video of multiple moving fixation points within a dynamic scene. According to one embodiment, the system includes a plurality of non-moving image capturing devices oriented around the scene such that the entire scene is substantially within the field of view of each image capturing device. The image capturing devices may be, for example, camera banks, each having a number of non-moving cameras, panoramic wide field of view cameras, or a combination thereof. Output from the image capturing devices is input to a number of image generators, one for each image capturing device. The image generators are capable, given a viewing angle and zoom parameter, of computing an image frame that a virtual camera positioned at its associated image capturing device, pointing to the given viewing angle and with the given zoom, would have output.

[0007] A first of the image generators is controlled by a control unit, which receives viewing angle and zoom parameter commands, such as from an operator via a user interface. These

viewing angle and zoom parameters are communicated to the first image generator. The appropriate viewing angle and zoom parameter commands for the remainder of the image generators are determined by a mapping module based on the viewing angle and zoom commands from the control unit and from data regarding the calibration between the various image capturing devices. The output image frames from certain of the image generators is input to an image sequence module, which outputs these images in sequence in the order of the placement of the image capturing devices around the scene as desired to generate the 3D surround-view image sequence.

[0008] The present invention provides numerous advantages in comparison to the relevant prior art. First, the system never misses an action of interest within the dynamic scene. In the previous master-slave pan/tilt-based systems, a human operator is tasked to identify and track a single action of interest, and all the cameras follow that action. Therefore, if (i) an action of true interest is occurring somewhere else within the scene, (ii) the operator's tracking is delayed, or (iii) the pan/tilt devices have servoing errors or delay, then the system will fail to capture video of the action, either totally or partially. In contrast, the system of the present invention favorably permits capturing all of the images in the scene all of the time.

[0009] Second, the present invention permits having multiple fixation points simultaneously. Third, the image generators need not track the target mechanically, and as such do not suffer from any control delay, offset, or other errors associated with servoing. Therefore, matching the point of rotation among images, as is described hereinafter, is improved in comparison with previous systems. Fourth, with the present invention the video can be replayed based on time (forward or backward), based on space (clockwise or counter-clockwise), or any combination thereof.

[0010] These and other benefits will be apparent from the description to follow.

BRIEF DESCRIPTION OF THE FIGURES

[0011] The present invention is described in conjunction with the following figures, wherein:

[0012] Figure 1 is a block diagram of a system for obtaining video of multiple moving fixation points within a dynamic scene according to one embodiment of the present invention;

[0013] Figure 2 is a diagram of an image capturing device of Figure 1 according to one embodiment of the present invention;

[0014] Figure 3 is a diagram illustrating the field of views (FOVs) of the cameras of the camera bank of Figure 2 according to one embodiment of the present invention;

[0015] Figure 4 is a diagram illustrating a situation in which the FOV of a virtual camera for a camera bank is within the FOV of a camera of that camera bank;

[0016] Figure 5 is a diagram illustrating a situation in which the FOV of a virtual camera for a camera bank is within the FOV of two cameras of that camera bank;

[0017] Figure 6 is a diagram of a portion of the system according to another embodiment of the present invention;

[0018] Figure 7 is a diagram of a portion of the system according to another embodiment of the present invention; and

[0019] Figure 8 is a diagram of the system according to another embodiment of the present invention.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0020] It is to be understood that the figures and descriptions of the following embodiments have been simplified to illustrate elements that are relevant for a clear understanding of the present invention, while eliminating, for purposes of clarity, other elements. For example, certain operating system details and modules of computer processing devices are not described herein. Those of ordinary skill in the art will recognize, however, that these and other elements may be desirable in a typical telecommunications device. However, because such elements are well known in the art, and because they do not facilitate a better understanding of the present invention, a discussion of such elements is not provided herein.

[0021] Figure 1 is a diagram of a system 10 for obtaining video of multiple moving fixation points within a dynamic scene 12 according to one embodiment of the present invention. The system 10 includes a number of static (i.e., non-moving) image capturing devices $14_{1\text{-}n}$ surrounding the scene 12. The field of view (FOV) of each of the image capturing devices $14_{1\text{-}n}$ may completely or substantially cover the entire scene 12.

[0022] As illustrated in Figure 2, according to one embodiment, the image capturing devices $14_{1\text{-}n}$ may include camera banks, each camera bank 14 including a number of static (i.e., non-moving) cameras $16_{a\text{-}i}$. Each camera $16_{a\text{-}i}$ may be fixated on a separate portion of the scene 12, as illustrated in Figure 3, such that the scene 12 is covered by at least one of the cameras $16_{a\text{-}i}$ of each camera bank $14_{1\text{-}n}$. That is, any point in the scene 12 is within the field of view of at least one of the cameras $16_{a\text{-}i}$ of each camera bank. In addition, the cameras $16_{a\text{-}i}$ on each particular bank may be aligned such that their imaging centers are substantially the same or as close as possible. According to one embodiment, the cameras $16_{a\text{-}i}$ are synchronized with a common signal so that the shutter for each camera $16_{a\text{-}i}$ fires at precisely the same time, resulting in video

frames taken at the same time instant. In addition, the timing of each video frame may be labeled electronically, i.e., time-stamped. Although nine cameras are shown in the camera bank illustrated in Figure 2, a different number of cameras may be used depending on the application.

[0023] According to other embodiments, one, some or all of the image capturing devices $14_{1\text{-}n}$ may be, for example, a single camera having a panoramic wide field of view. Such a panoramic wide FOV camera may include, for example, a parabolic or spherical mirror with which a wide angle of view is mapped to the imaging surface of the camera. According to another embodiment, the panoramic wide FOV camera may include, for example, a fish-eye lens with which the wide angle view is captured onto the imaging surface of the camera.

[0024] The number of image capturing devices 14 included in the system 10 may depend, for example, on the desired quality of the 3D surround-view image sequence to be generated. For example, the system 10 may include ten to eighty image capturing devices 14 surrounding the scene 12. In addition, according to another embodiment, the image capturing devices 14 may be periodically positioned around the scene 12 such as, for example, every five degrees or every ten degrees.

[0025] Returning to Figure 1, for an embodiment in which the image capturing devices 14 are camera banks such as illustrated in Figure 2, the system 10 may also include a video multiplexer $18_{1\text{-}n}$ coupled to each of the image capturing devices $14_{1\text{-}n}$. The output from each of the cameras $16_{a\text{-}i}$ may be multiplexed onto, for example, one video fiber cable by the video multiplexer $18_{1\text{-}n}$ that feeds to a separate image generator $20_{1\text{-}n}$ and video storage unit $22_{1\text{-}n}$ for each image capturing device $14_{1\text{-}n}$. The output video of the cameras $16_{a\text{-}i}$ may be digitally stored on a continuous basis in the respective video storage units $22_{1\text{-}n}$, enabling fast retrieval of corresponding frames in time for all cameras. For an embodiment in which the image capturing

device 14 is a panoramic wide FOV camera, the video multiplexer coupled between the camera and the image generator may be eliminated.

[0026] The image generators 20_{1-n} may be implemented as computers, such as workstations or personal computers, having software which when executed cause the image generators 20_{1-n} to compute an image frame that a virtual camera at its respective image capturing device 14_{1-n} would have output, given a particular viewing angle and zoom parameter. The process consists of first backprojecting the image frame of the virtual camera onto the scene 12 to obtain the field of view (FOV) of the virtual camera. Where the image capturing devices 14 include camera banks, if the FOV is completely within the field of view of one of the real cameras 16_{a-i}, as illustrated in Figure 4, then the virtual image may be obtained by cropping the corresponding region from the real image and transforming it perspectively (or simply scaling the image may suffice). If the virtual FOV overlaps across the FOVs of two or more real cameras 16_{a-i}, as illustrated in Figure 5, the virtual image may be obtained by cropping the part images, transforming each of the part images by an appropriate perspective transformation that corresponds to the transformation from the imaging plane of each physical camera 16 to that of the virtual camera, and finally merging the transformed part images into a single frame. The process is known as panoramic mosaicing, and is described generally in H. Shum and R. Szeliski, "Construction of Panoramic Mosaics with Global and Local Alignment," International Journal of Computer Vision, Vol. 36(2): 101-130, Feb. 2000, which is incorporated herein by reference. To account for the calibration between the cameras 16_{a-i} on their particular camera bank 14_{1-n}, each image generator 20_{1-n} may include an intra-bank calibration database 26_{1-n} that contains data regarding the calibration between the cameras 16_{a-i}. The intra-bank calibration database can be created by calibrating the intrinsic parameters of each camera and the relative

pose between cameras using, for example, well-known camera calibration algorithms used in the fields of computer vision and photogrammetry. For embodiments in which an image capturing device 14 is a panoramic wide FOV camera, the intra-bank calibration database for the corresponding image generator 20_{1-n} may be eliminated.

[0027] Where an image capturing device 14 is a panoramic wide field of view camera with a mirror, the corresponding image generator 20 may first crop the part of the image corresponding to the space angle of the FOV of the virtual camera and then transform the cropped image to remove the distortion that is contained due to the mirror, parabolic, spherical or otherwise. Where the image capturing device 14 is a camera with a lens, the corresponding image generator 20 may first crop the part of the image corresponding to the space angle of the FOV of the virtual camera and then transform the cropped image to remove distortion that is contained due to the lens, fish-eye or otherwise.

[0028] For replay operation, the virtual image generators 20_{1-n} may retrieve the video data stored in the video storage units 22_{1-n}. For real-time operation, the virtual image generators 20_{1-n} may use the real-time video from the image capturing devices 14_{1-n} as it is stored in the storage units 22_{1-n}.

[0029] One of the image generators 20_{1-n} may serve as a master virtual camera. For purposes of this discussion, assume the master virtual camera is the image generator 20_1. An operator of a control unit 24 may control the view seen by the master virtual camera based on viewing angle and zoom parameter commands input to the image generator 20_1. The control unit 24 may include an operator interface such as, for example, a pointing device plus a video display. According to one embodiment, the control unit 24 may be similar to a traditional cameraman's tripod, with angle sensors plus zoom and height control knobs. The video seen by the operator

on the display monitor would be the video generated by the image generator 20_1. According to another embodiment, the control unit 24 may be a computer terminal where, for example, a mouse or other type of input device is used to input the viewing angle and zoom parameter commands. The control unit 24 interprets the desired pan/tilt angle and zoom from the operator's commands, and inputs them to the image generator 20_1 of the virtual master camera.

[0030] As illustrated in Figure 1, the system 10 also includes a surround-view image sequence generator 30. The surround-view image sequence generator 30 includes an image capturing device mapping module 32 and an image sequencing module 34, and has an associated inter-image capturing device calibration database 36. The generator 30 may be implemented on a computer such as, for example, a workstation or a personal computer. The modules 32, 34 may be implemented as software code to be executed by the generator 30 using any suitable computer language such as, for example, Java, C or C++ using, for example, conventional or object-oriented techniques. The software code may be stored as a series of instructions or commands on a computer readable medium, such as a random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard-drive or a floppy disk, or an optical medium such as a CD-ROM. According to one embodiment, the modules 32, 34 may reside on separate physical devices.

[0031] The mapping module 32 may receive the viewing angle and zoom parameter commands from the control unit 24 and, based thereon, may compute the three-dimensional location of the action of interest in the scene 12. Using calibration data regarding the calibration between each image capturing device $14_{1\text{-}n}$, which is stored in the inter-image capturing device calibration database 36, the mapping module 32 computes the corresponding viewing angles and zoom data for the virtual cameras of the other image generators $20_{2\text{-}n}$.

[0032] The output from certain (all or less than all) of the image generators $20_{1\text{-}n}$ is supplied to the image sequencing module 34, which may output these images in sequence in the order of the placement of their image capturing devices $14_{1\text{-}n}$ around the scene 12, either clockwise or counter-clockwise, to generate the 3D surround-view image sequence.

[0033] As discussed previously, the inter-image capturing device calibration database 36 stores data on the relationship between each image capturing device 14 to the scene 12 and to the other image capturing devices. This data may be determined prior to operation of the system. According to one embodiment, appropriate calibration requires determining the pose (location and orientation) of each of the image capturing devices $14_{1\text{-}n}$ with respect to a scene coordinate system. In addition, it includes determining the relationship of the zoom control parameter (from the control unit 24) to angular field of view, and determining the relationship of the focus control parameter (from the control unit 24) to the distance of objects in the scene 12.

[0034] Image capturing devices 14 pose may be determined, according to one embodiment, by determining the proper viewing angle of the image generators $20_{1\text{-}n}$ for a set of distinguished points or "landmarks" with known 3D coordinates. The viewing angle parameters may be stored with the (x,y,z) coordinates of the landmark to form one pose calibration measurement. According to another embodiment, camera bank pose may be determined by an optimization procedure, using three or more landmark measurements in a nondegenerate configuration.

[0035] The mapping module 32 computes the corresponding viewing angles and zoom data for the virtual cameras of the other image generators $20_{2\text{-}n}$ based on the viewing angle and zoom parameter commands from the control unit 24. According to one embodiment, this may be performed by first determining the equation of a 3D line specifying the principal viewing ray of the virtual camera of the first image generator 20_1. All points on this line can be represented as

$p = c + kv$, where p is a 3D point on the line, c is the focal point of the virtual master camera of image generator 20_1, and v is a unit vector representing the orientation of the principal axis, directed out from the focal point, and k is a scalar parameter that selects different points on the line. Only points on the line that are in front of the focal point (i.e., $k > 0$) are considered to be on the master camera principal viewing ray.

[0036] The desired virtual servo-fixation point (VSFP) for the surround-view effect is defined to be some point on the principal viewing ray of the master virtual camera of the image generator 20_1. Choosing which point is the VSFP is equivalent to choosing a value for parameter k in the above-equation. The VSFP can be determined by intersecting the principal viewing ray with an equation or set of equations representing, for example, a real surface in the scene 12, a virtual (nonphysical) surface in the scene 12, or a combination thereof. If there is more than one intersection point, the desired VSFP should be determined. According to one embodiment, the point closest to the camera bank 14_1 is chosen. In addition, if there is no mathematical intersection point, an alternate method may be used to determine the VSFP. According to one embodiment, the last known valid point of intersection is used.

[0037] For each of the other image generators $20_{2\text{-}n}$, the mapping module 32 uses the 3D position of the VSFP to compute the viewing angle value that brings the virtual camera for each of the image generators $20_{2\text{-}n}$ principal-viewing ray into alignment with the VSFP. To calculate the zoom parameters for the image generator 20_i, where $2 \leq i \leq n$, the distance d between the position of the image capturing device 14_i and the VSFP is computed. If y is the position of the image capturing device 14_i and x is the VSFP, and vector $(a,b,c) = x\text{-}y$, then d may be computed as $d = \sqrt{a^2 + b^2 + c^2}$.

[0038] The zoom of each of the virtual cameras of image generators $20_{1\text{-}n}$ may be controlled to keep the point of interest the same size in all the images, even though the image capturing devices $14_{1\text{-}n}$ are different distances away from the object. Let r be the desired radius of a virtual sphere subtending the entire vertical field of view of each image. Let d_i be the distance from image capturing device 14_i to the VSFP. Then the desired vertical field of view angle α_i can be computed as $\alpha_i = 2 \cdot \arctan(r / d_i)$. The zoom parameter that achieves this desired field of view is then computed from data collected during the prior zoom calibration procedure.

[0039] To control the focus of each of the virtual cameras of the image generators $20_{1\text{-}n}$ to achieve sharp focus at the VSFP, the focus parameter that achieves sharp focus at distance d_i may be computed for image generator 20_i using the distance vs. focus parameters equations or tables derived from the prior focus camera calibration procedure.

[0040] Having computed the proper viewing angle and zoom parameter for each of the image generators $20_{2\text{-}n}$, the mapping module 32 communicates these values to the image generators $20_{2\text{-}n}$. The image generators $20_{2\text{-}n}$ then generate the appropriate image frame based on the video frames from the cameras $16_{a\text{-}i}$ of their respective camera bank $14_{2\text{-}n}$ (or from the image captured by a panoramic wide field of view camera). The images from each of the image generators $20_{1\text{-}n}$ is supplied to the image sequencing module 34, which outputs these images in sequence in the order of the placement of their image capturing devices $14_{1\text{-}n}$ around the scene 12, either clockwise or counter-clockwise, to generate the 3D surround-view image sequence.

[0041] As discussed previously, the control unit 24 outputs viewing angle and zoom commands to the master image generator. In addition, the viewing angle and zoom commands may be generated by a human operator via user interface as discussed previously. According to another embodiment, the human operator may be replaced with a computer vision module 40 as

illustrated in Figure 6. The computer vision module 40 may automatically detect and track moving objects (e.g., fixation points) within the scene 12 by processing video from the master image generator. The computer vision module 40 may be implemented as software code to be executed by a computer processing device, such as a workstation or a personal computer, using any suitable computer language such as, for example, Java, C or C++ using, for example, conventional or object-oriented techniques. The software code may be stored as a series of instructions or commands on a computer readable medium, such as a random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard-drive or a floppy disk, or an optical medium such as a CD-ROM. In addition, according to another embodiment, the computer vision module 40 may be able to automatically select a different image generator $20_{1\text{-}n}$ as the master image generator to, for example, decrease the distance between the image capturing device 14 corresponding to the master image generator and the object being tracked or to increase the visibility of the object being tracked.

[0042] In addition, according to another embodiment of the present invention, multiple master image generators may simultaneously co-exist. According to such an embodiment, multiple and distinct control units 24 may control a master image generator, as illustrated in Figure 7. As such, individual viewers of the surround-view image sequence may control the fixation point. According to another embodiment, separate control units 24 may control the same image generator 20 as separate master image generators.

[0043] According to other embodiments of the present invention, one or more servo-controlled, moving (e.g., pan/tilt) cameras 42 may be positioned around scene 12, as illustrated in Figure 8. Each pan/tilt camera 42 may have an image generator 20 associated therewith, as described previously. In a manner similar to that described previously, the pan/tilt cameras 42

may receive the viewing angle and zoom commands based on the output from the control unit 24. The field of view of the pan/tilt cameras 42 need not span the entire scene 12, but may be greater than the field of view necessary to capture images of the fixation point. Thus, the servo errors associated with the pan/tilt cameras 42 may be compensated for by computing the virtual video at their associated image generator 20 that would correspond to the case with no error, thereby realizing smoother transition of view points in the surround-view image sequence. Techniques for computing the virtual video for a pan/tilt camera that corresponds to the case of no servo error are described in provisional U.S. Provisional Patent Applications Serial No. 60/268,205 and Serial No. 60/268, 206, both filed on February 12, 2001 and both incorporated herein by reference. Such an embodiment may be advantageous for areas of high interest within the scene 12 because pan/tilt cameras typically have greater resolution than panoramic wide field of view cameras.

[0044] Embodiments of the present invention offer important and critical advantages over previous master-slave pan/tilt-based systems. First, one embodiment of the system never misses an action of interest within the dynamic scene 12. In the previous master-slave pan/tilt-based systems, a human operator is tasked to identify and track a single action of interest, and all the cameras follow that action. Therefore, if (i) an action of true interest is occurring somewhere else within the scene 12, (ii) the operator's tracking is delayed, or (iii) the pan/tilt devices have servoing errors or delay, then the system will fail to capture video of the action totally or partially. In contrast, the system of the present invention captures all of the images in the scene 12 all of the time.

[0045] Second, one embodiment of the present invention permits having multiple fixation points simultaneously. Third, in one embodiment the so-called slave virtual cameras of image

generators 20_{2-n} do not track the target mechanically, and as such will not suffer from any control delay, offset, or other errors associated with servoing. Therefore, matching the point of rotation among images, as is described hereinafter, is improved in comparison with previous systems. Fourth, with one embodiment of the present invention the video can be replayed based on time (forward or backward), based on space (clockwise or counter-clockwise), or any combination thereof.

[0046] Although embodiments the present invention has been described herein with respect to certain embodiments, those of ordinary skill in the art will recognize that many modifications and variations of the present invention may be implemented. The foregoing description and the following claims are intended to cover all such modifications and variations.